KATS: Kernel-smoothed Adaptive Thresholds for Network Anomaly Detection

Joshua Neil, Alexander Brugh, ACS-PO;
Curtis Storlie, CCS-6; Michael Fisk, ACS-PO

We are interested in detecting anomalous activity in computer networks. Our approach is to monitor counts of network activity per unit time. To monitor these counts, we have designed KATS, which uses a baseline model of activity and examines deviations from this baseline. The model captures daily and weekly trends and is allowed to change over time, as the data suggests. Anomalies are accumulated over time so that both a single highly anomalous count and/or a sequence of moderately anomalous counts will trigger alarms in the system.

Network anomaly detection at LANL has evolved over the past few years. It began with a system known as Exponentially Weighted Moving Average Anomaly Detection (EMAAD), which used exponentially weighted moving estimates of the mean and variance, and a threshold based upon these estimates, to alarm when network counts became too large. One major problem with this approach is that it has no awareness of daily/weekly trends in the data. Therefore, thresholds had to be set very high to avoid false alarms, and detection of true anomalies was nearly impossible. To replace EMAAD, we have developed Kernel-smoothed Adaptive Thresholds (KATS), which we describe here. KATS uses the same data stream that EMAAD uses. To keep the description brief, we will not go into the details of EMAAD, but simply point out a few advantages our system holds over it:

- **A better model** – Empirical results show that while EMAAD assumes a Gaussian distribution for the data, the data is more accurately represented by a negative binomial distribution.
- **A better way of handling daily and weekly periods in the data** – This will be described in more detail under the description of the algorithm that follows.
- **A method for accumulating anomalies** – This approach uses control chart theory, and is described under the description of the algorithm that follows.
- **A more sophisticated thresholding mechanism** – This increases the detection power of the approach.

KATS is loosely based upon the technique described in [1]. In this paper we describe this algorithm, followed by the implementation details and initial results. Finally, we describe future work. We note that while the basic ideas in this paper have proven effective, the modifications required to implement this technique were extensive, and much remains to be gained from improvements to this system.

The system described by Lambert and Liu [1] works by using a model to estimate the probability of observing a given count, followed by thresholding a severity score based on this probability. If we observe a count that is too rare according to the model, we sound an alarm.

The basic model for any given time t is a negative binomial model, with mean and variance . One immediate conclusion one can make from examining network data is that it tends to be periodic in the time of day, and also the day of the week. Monday at 4:00 am looks quite different than Monday at 8:00 am. Also, Thursday looks different than Friday, which looks different than Saturday or Sunday. To model the differences within a day, we have a probability function parameterized for each minute in the day. The parameters for the probability function are estimated on a course grid across the day. Between the grid points, the parameters are estimated using kernel smoothing of the grid points. Currently, we have grid points for each 10 minutes of the day. When we receive a count at some minute t, we will form a locally weighted average using the 10-minute grid points to estimate μ_t, and we form another smoothed curve to estimate σ^2_t. The grid points are updated using an exponentially weighted moving average (EWMA) from previous grid points at that time, so that we get a smooth, accurate picture of the normal activity for that grid point. Finally, to account for differences

For more information contact Joshua Neil at jneil@lanl.gov.

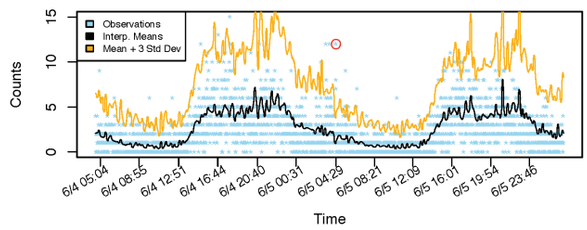

Fig. 1. Counts for one host.

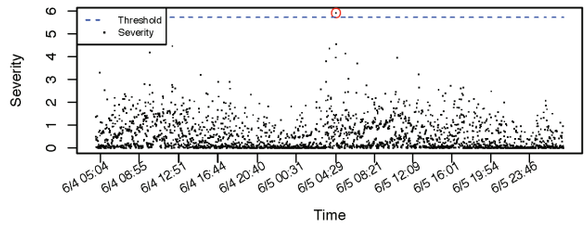

Fig. 2. Severity metric.

across the week, we maintain three sets of grid points, one set for Monday through Thursday, another for Friday, and the third for Saturday and Sunday.

When a count C_t is observed at time t, we measure its p-value, the probability of seeing something as large or larger than that C_t, according to the model at time t. These p-values should be approximately uniformly distributed on the interval (0, 1). Thus, evaluating the inverse cumulative distribution function of a standard normal distribution at that p-value should result in an approximately standard normal distribution. These so-called Z scores are accumulated using a cumulative-sum approach to form a severity metric, S_t. This provides for anomaly detection that can react to either an extremely rare count or a succession of moderately rare counts.

Figure 1 shows an example of the algorithm run on 2 days of data collected from LANL's yellow network. The top graph shows per-minute counts of session activity along with mean and variance estimates. An observation that triggered an alarm is circled in red. The lower graph shows the corresponding severity metric and also features a point circled in red. Several high counts can be observed prior to the alert, which contributed to a building severity value that eventually exceeded the threshold at which we alarm.

We list several aspects of this work that can be easily improved upon:

- **Tuning the Threshold and EWMA Weights** – Currently, we are tuning the threshold parameters by hand. A more accurate approach will involve using the normality of the severity metric to set appropriate threshold values, yielding desired false-alarm rates. We anticipate some adjustment as needed to achieve acceptable system performance. In addition, the weights on the EWMAs may need adjustment.
- **The Negative Binomial Model** – There is evidence that the negative binomial model is not flexible enough to handle the variety of user behavior we see in the network. There are two major drawbacks to this model: (1) the mean must be smaller than the variance; so larger counts must therefore correspond to higher variability, a fact not reflected in all data sets, and (2) the negative binomial can only represent data with positive (right) skew, and we have observed counts with negative skew. We are considering a discretized Weibull distribution, which fixes both of these problems.
- **Maximum Likelihood Estimates** – Currently, the grid means and variances are based on method of moments estimators. More accurate results may be achieved by using maximum likelihood estimates.

[1] Lambert, D. and C. Liu, *J Am Stat Assoc* **101**, 78 (2006).

Funding Acknowledgments
Site Cyber-Security Initiatives